RESEARCH ARTICLE OPEN ACCESS

Performance of data mining algorithms in unauthorized intrusion detection systems in computer networks

Hadi Ghadimkhani, Ali Habiboghli*, Rouhollah Mostafaei
Department of Computer Science and Engineering, Islamic Azad University, khoy Branch, Khoy, Iran
Corresponding Author : Ali Habiboghli

--------------------------------------********************************---------------------------------

Abstract:

In this article, we first normalized data by using random sampling conducted and chose a limited number of data. Then we determined the two types of clusters using hours and non-working hours and hours, and we implement the clusters ID3 and Decision tree algorithms. The results of our proposed method show that ID3 algorithm for high power at sorting and classification acts better than Decision tree. The purpose of this paper is to compare the performance of these two algorithms.

Keywords **—Intrusion Detection, Network, Data Mining.**

--------------------------------------********************************---------------------------------

I. INTRODUCTION

In today's world, computers and computer networks connected to the Internet will play a major role in communication and information transfer. In the meantime, jobber with access to important data or the information centres of other people and with the intent to influence or pressure, or even to disassemble order systems, have adopted the act to the computer systems. Thus the need to protect the information security and maintain the effectiveness in computer networks that communicate with the outside world is necessary [7].

An Intrusion Detection System (IDS) is a security technique attempting to detect various attacks. It has been identified mainly two techniques, namely misuse detection and anomaly detection [1, 2, 8, 9].

With the growth of information technology, network security is proposed as an important issue and very large challenge [7]. Intrusion detection system is the main component of a secure network. Traditional intrusion detection systems cannot adapt with new attacks, therefore Data mining-based intrusion detection systems offered today. Here we use data mining methods and reduced features and Clustering and classification techniques on our database.

Misuse detection, also called signature-based detection, attempts to model abnormal behaviour and normally focuses on the known attacks. It uses a descriptive language to delineate the characteristics of the known attacks and to construct the corresponding attack signatures. However, it may not be able to alert the system administrator in case of a new attack. Anomaly detection attempts to model normal behaviour. Any events which deviates the normal usage patterns are considered to be suspicious. It constructs the profile of user behaviour or status of network traffic and compares the observed behaviour with the stored profile to determine whether an attack action occurs. The anomaly detectionapproach may have the advantage of detecting previously unknown attacks over the misuse detection approach. However, it may suffer from false alarm problem and radically changed user behaviours [3].

Except the above taxonomy of intrusion detection techniques, IDS is classified into host-based and network-based IDS by their defensive scopes [10].

II. DESCRIPTION OF PROBLEM

A. Network intrusion detection system

A network intrusion detection system or Network-Based IDS, Briefly to say NIDS, In fact, is a variety of network intrusion detection system [4] which is connected to the network and in this way, monitoring network traffic and provides reports. Methods of making such systems are usually behind or in front of the firewall network. In Figure 1you can see how to position the network intrusion detection system. Always the best location for a network intrusion detection system NIDS, is outside the network firewall.

Jing and Papavassiliou [5] propose a new network traffic prediction methodology based on the frequency domain traffic analysis filtering.

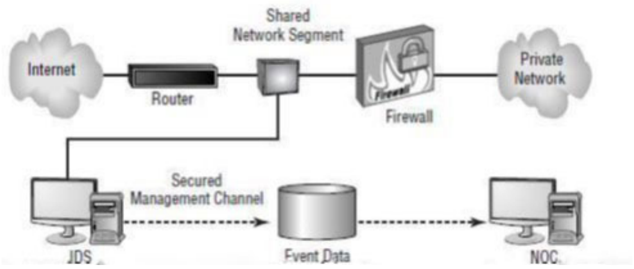


Figure 1. Wrapping method NIDS in network shows how to analyse network traffic.

B. *Architectural types of intrusion detection systems*

Various architectural intrusion detection systems include:

- Host-based intrusion detection system (HIDS)
- Network-based intrusion detection system (NIDS)
- Distributed Intrusion Detection System (DIDS) [6].

B.1. Host-based intrusion detection system (HIDS)

The system is responsible to detect unauthorized activity on the host computer .Host-based intrusion detection system can detect attacks and threats to critical systems (Includes access to files, Trojans and etc.) that are not detectable by the network-based intrusion detection systems. HIDS due to their location on the host to be monitored are notified all types of additional local information by implement security (Including system calls, change system files and system connections).

This when combined with the communications network, may provide a good data to search for possible event [6].

B.2. Network-based intrusion detection system (NIDS)

NIDS name is derived from the fact that of where it is located, the entire network is monitored.

Network-based intrusion detection systems adapt to detect unauthorized intrusions before they reach critical systems. NIDS are mostly formed two monitors (sensors) and factor. These are often installed behind the firewall and other access points to detect any unauthorized activity [6].

B.3. Distributed Intrusion Detection System (DIDS)

These systems are composed of several HIDS or NIDS or the combination of these two with a central management station.

The statements that any IDS are available on the network send their reports to the central management station. Central Station is responsible to examining the reports and alert security officer. This central station also is responsible to updates detection rule base of each IDS on the network [6].

III. PROPOSED METHOD

For network analysis, a method used that to the help the centrality of the concept of dependency, directed graph become hierarchical graph. Using hierarchical structure can be easily distinguished between leaders and followers till the terrorist network to be built. New ideas for measuring the centrality dependence is much to do hierarchical model, for this idea shows nodes that are entirely dependent on a particular node. How to calculate the central of attachment is in Formula 1[11, 12].

The central of attachment

$$DC_m = \sum_{m\neq p,p\in G} \frac{d_{mn}}{N_p} + \Omega \qquad (1)$$

That

$N_p =$ Number of nodes per cluster

And

$m, n =$ The distance between two nodes

Using measurements of network performance is well represented the effect of each node of the graph. The removal of a node, the further decrease network performance ,the importance of their presence is tied more. The calculation of network performance is displayed in formula 2.
Formula 2, the d_{ij} value represents the shortest distance between two nodes of j and I [11, 12].

Efficiency centrality

$$E(G) = \frac{\sum_{i \neq j} e_{ij}}{N(N-1)} = \frac{1}{N(N-1)} \sum_{i \neq j \epsilon G} \frac{1}{d_{ij}} \quad (2)$$

That

$i, j =$ The distance between the node

And

$N \ = \ Number\ of\ nodes$

And

$d_{i,j} =$ The shortest distance between two nodes

Central Profile of organizational role specifies the role of the network. Corporate roles Profile will be counted by using centralized network performance. If the central organizational role profile, plotted on the coordinates, the top of the axis represents leaders and below the horizontal axis, represents the X role of soldiers and less important members of the group.
How to calculate the profile of organizational role are shown in Formula 3 [11,12].

Organizational designation index

$$PRI = E(G) - E(G - node_i), i = 1,2, \dots, N \quad (3)$$

That

$E(G) = The\ centrality\ of\ the\ total\ performance$

And

$node_i = Central\ node\ performance$

IV. EXPERIMENTAL RESULTS

A. *Selection step and data collection*

Here we use a limited number of data which are includes features as date, time, user, pc, activity.

B. *Data preparation*

At this stage, first discrete data is carried out.

Activity(count)	activity	date	row
115	logon	2010/04/01	**1**
28	logoff	2010/04/01	**2**
34	logon	2010/05/01	**3**
34	logoff	2010/04/01	**4**

Table 1- the number of log off, log on by date

C. *Sampling*

Here, first we have Categories the data based on date and connect and disconnect rate:
So, as you can observe according to table1 and figure 2 (2010/04/01), numbers of logon are more than logoff and it can be cited that this date can be mystification for us.

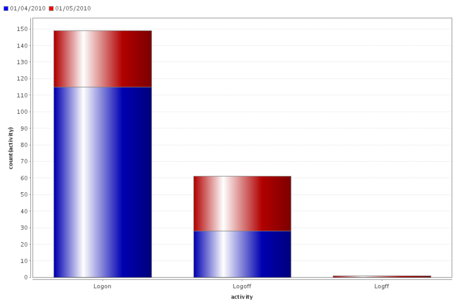
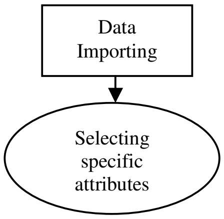


Figure 2. Frequency of logon and logoff according time

In Figure 3 you can see the ID3 algorithm implementation process.



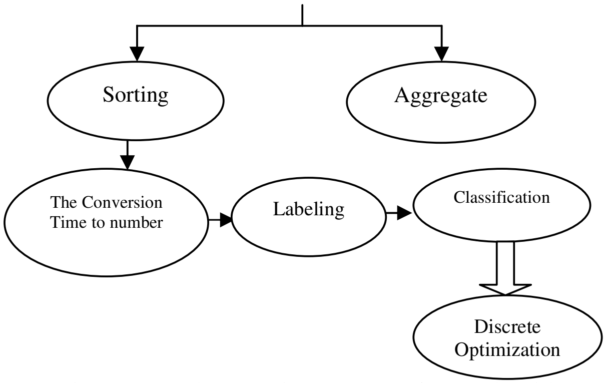


Figure 3. Implementation Process of ID3 Algorithm

Decision tree algorithm implementation process can be seen in Figure 4.

File CSV is the study input that was taken from a site that was said and enters into Rapid Miner software for analysis.

According table 2, we have turn the clocks in number to be able to do calculations more comfortable. For this aim, first we have sorting the date and time ascending.

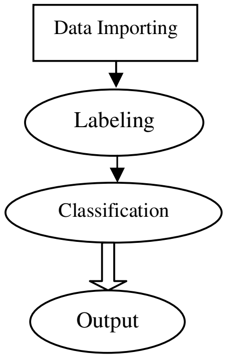
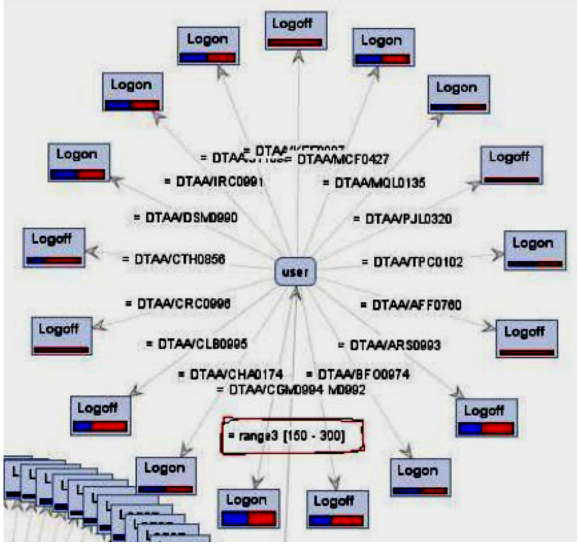


Figure 4. Implementation Process of Decision Tree Algorithm

Numerical Conversion	Time	date	row
0	07:00:10	2010/04/01	1
1	07:00:53	2010/04/01	2
2	07:01:09	2010/04/01	3
3	07:01:52	2010/04/01	4
4	07:02:27	2010/04/01	5
5	07:03:13	2010/04/01	6
6	07:03:17	2010/04/01	7
7	07:03:23	2010/04/01	8
8	07:03:52	2010/04/01	9
9	07:04:05	2010/04/01	10

Table 2-Conversion time to number and sort them ascending

Figure 5-Clustering Chart id3



Thus, as you can infer according to figure 5, first hours of day , all users are logon, but in range [150-300] that is related to last hours of work , it can be cited that, all of users has gotten logon same as logoff hours.

But, you can understand with care that DTAA/CRC0996 - DTAA/KEE0997-DTAA/AFF0760 users are logoff perfectly in these work hours, for more accurate security , manager can pay attention to clustering algorithm written part that as sample has been displayed in figure 6.

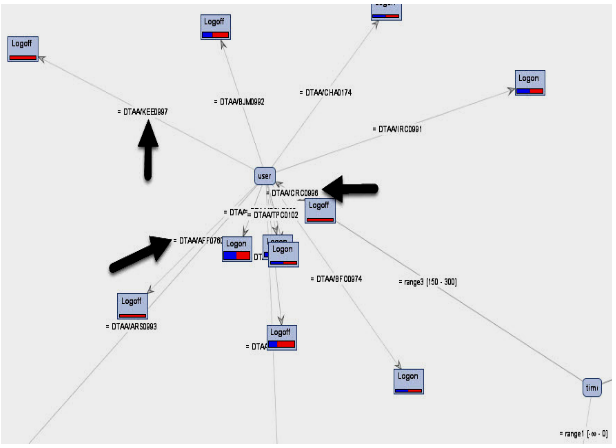


Figure 6. graph clustering of ID3

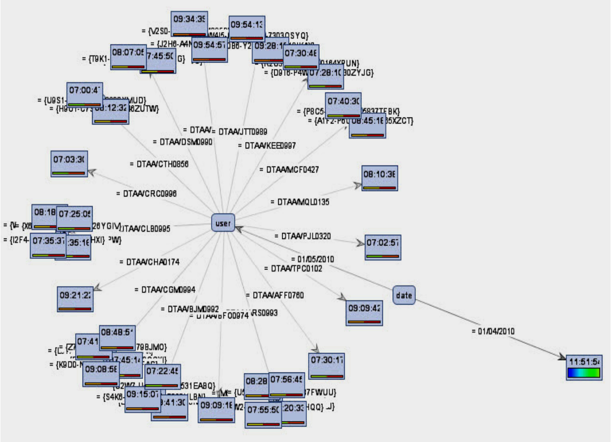


Figure 7- Decision tree diagram

As you can see in Figure 7, in this algorithm can be seen different id to a particular user on 01/05/2010 but this algorithm has not done well classification to date 01.04.2010 that could be a weak point of the algorithms. If you stand on any of the classes built, you can see the members of the class and through this comparison, we can say that the classification ID3 is much better than decision tree.

V. CONCLUSION

To establish complete security in a computer system, in addition to firewalls and other intrusion prevention systems called intrusion detection systems (IDS) are needed to see if the attacker of firewall, antivirus, other security was passed and signed, it detects and think of ways to deal with it. In this study, we used data information network to possible data mining and intrusion detection or any attack or any unacceptable action which was conducted according to regulations and laws.We first normalized data analysed using random sampling and volume data on 300 others. Then ID3 and decision tree algorithm on them was implemented. ID3 algorithm acts better than decision tree in packing and sorting, and this is one of the strengths of ID3 algorithm, however, Decision Tree cannot sorted the information, but in processing acts better than ID3. Decision tree algorithm has high error probability in the case of low and high category, Because of the limited number of data used in this article.

REFERENCES

[1].C. M. Chen, Y. L. Chen and H. C. Lin, "*An efficient network intrusion detection*", Computer Communications (33), pp: 411–424, 2111.

[2].T. Verwoerd, R. Hunt, Intrusion detection techniques and approaches, Computer Communications 25 (15) (2002) 1356–1365.

[3].E.Lundin and E.Jonsson, "Anomaly-based intrusion detection: privacy concerns and Otherproblems", Journal of Computer Networks, volume (3), number (4), pp: 623-640،2111.

[4].P. Dokas, L. Ertoz, V. Kumar, A. Lazarevic, J. Srivastava and P. N. Tan, "Data Mining For Network Intrusion Detection", Computer Science Department, 211 Union.Street SE,4-102, EE/CSC Building University of Minnesota, Minneapolis, MN 11411, USA, 2012.

[5].J. Jing, S. Papavassiliou, Enhancing network traffic prediction and anomaly detection via statistical network traffic separation and combination strategies, Computer Communications 29 (10) (2006) 1627–1638.

[6].EhsanMalekian, An attacker on the network and coping strategies, Nas Publisher, 2009, in Persian

[7]. Masoud Sotoudeh Far, Network intrusion detection based on adaptive fuzzy rules,master thesis, 2004, in Persian.

[8]. E. Biermann, E. Cloete, L.M. Venter, A comparison of intrusion detection systems, Computer and Security 20 (8) (2001) 676–683.

[9]. J.M. Estevez-Tapiador, P. Garcia-Teodoro, J.E. Diaz-Verdejo, Anomaly detection methods in wired networks: a survey and taxonomy, Computer Communications 27 (16) (2004) 1569–1584.

[10]. H.S. Venter, J.H.P. Eloff, A taxonomy for information security

technologies, Computer and Security 22 (4) (2003) 299–307.

[11]. Memon Nasrullah and Henrik Legind Larsen, Practical Approaches for Analysis, Visualization andDestabilizing Terrorist Networks. In the proceedings ofARES 2006: The First International Conference onAvailability, Reliability and Security, Vienna, Austria,IEEE Computer Society, pp. 906-913,2006.

[12]. MemonNasrullah and Henrik Legind Larsen, Practical Algorithms of Destabilizing Terrorist Networks. In the proceedings of IEEE Intelligence Security Conference, San Diego, Lecture Notes in Computer Science, Springer-Verlag, Vol. 3976: pp. 398-411 , 2006.